

ASSIGNMENT #2: MISLEADING VISUALIZATION

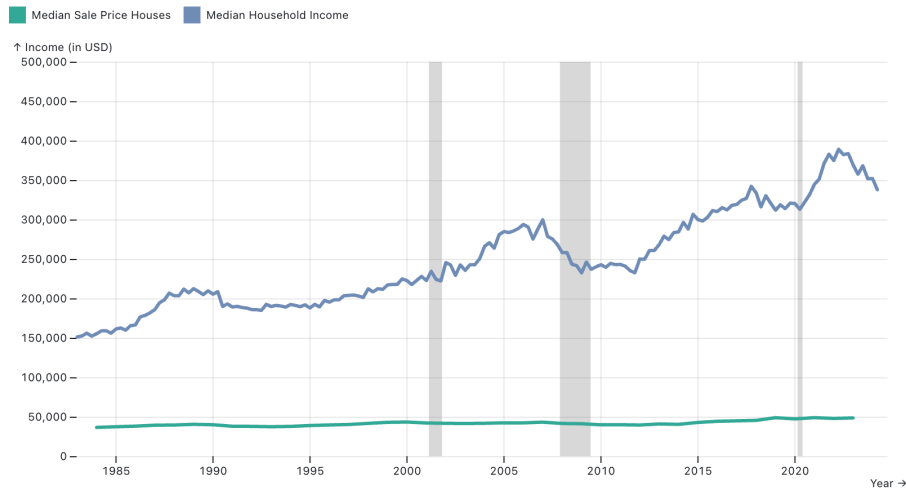
Ivan Aristy — iae225

CS-GY 6313 B

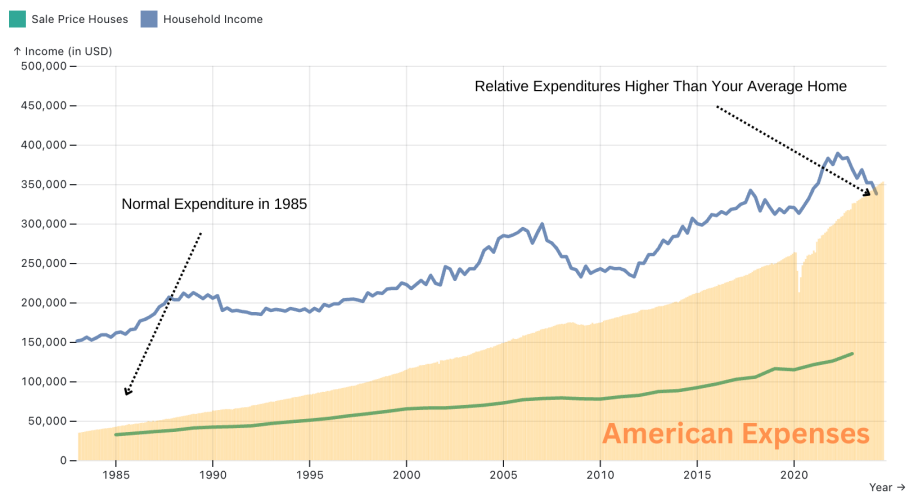
1 Visualizations

The following are the two Visualizations that I have created for this Assignment:

**Real Median Home Sale Price vs Real Median Household Income: US
(January 1st 2000 as reference year)**



Home Prices vs Household Income



2 Report

2.1 Topic Selection

2.1.1 Topic Ideation

For the Misleading visualization I had a few ideas:

- Immigration and Crime Correlations: Showing an extrapolated correlation between immigration and crime rates could be made misleading by selecting data points that support a biased narrative. For example, not filtering by region, and actively sampling from regions with high crime rates.
- Ethnically Biased Drug Usage: Using the "US Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin" [2] dataset, we could use misleading bar charts to exaggerate the correlation between hispanic heritage and drug consumption. Additionally, by filtering out specific drugs we could make it especially misleading.
- Presidential Employment Claims: As much as I do not like Trump, which you'll see in the extra credit, the high employment rates during his presidency are largely do to the Covid-19 pandemic. We could use a line chart to show the employment rates over time, in contrast/comparison to Biden, and by not clarifying the context of the data, we could make it seem like Trump was solely responsible for the high employment rates.
- Immigration Crisis and Democratic Party Correlation: We could use a line graph to show the number of border crossings over time, and, by comparing Biden's and Trump's presidency, we could make it seem like the Democratic party is solely responsible for the immigration crisis. This could be further exaggerated by not showing the Covid-19 pandemic context, and/or picking legal border crossings to be also part of the data using the "Border Crossing Entry Data" [1] dataset.
- The Cost of Living vs Inflation: The relation between the American people's expenditures, home prices, income, and inflation rates are great for generating accurate and misleading content alike. For example, one could greatly overvalue the spending power of the American people by not accounting for inflation rates. Also, by sampling subsets of the population, we could make homes seem less or more affordable than they are.

2.1.2 Topic Selection and Question

I decided to go with the "The Cost of Living vs Inflation" topic.

The reason why is because cost of living and income statistics lend themselves to both areas of the spectrum of honesty. We can sample and transform numbers in multiple ways, incorrectly scale items, and/or dramatize the effects of specific economic factors to create misleading narratives on the issues Americans care most about. This topic also facilitates my secondary goal for this assignment: Showing how changes in data selection and visualization techniques affect the perception of the viewer.

The question to answer is: *Has the increase in home prices made it harder for the average American to own a home?*

I believe that this question will be perfect for evaluating both ends of the spectrum.

We will first try to generate the correct visualization of the data, answering the question in the most clear and attentive way possible. Then, I aim to generate a visualization that creates confusion through various mediums. For example, the misleading visualization downplays, exaggerates, or eliminates important considerations like inflation, and/or misrepresents the increase in home prices through incorrect sampling.

2.2 Data

2.2.1 Required Data

We will focus on the analysis of yearly data.

The goal was to acquire data regarding:

- Home Prices: Real Home Prices, Non-Inflation Adjusted Home Prices, Home Price Index data, Rental Prices, Mortgage Prices, or Mortgage Rates.
- Income: Real Income, Non-Inflation Adjusted Income, Income Index data, or household earnings.
- miscellaneous: American spending habits and/or miscellaneous expenditures.

2.2.2 Data Gathering

All data was obtained from the Federal Reserve Economic Data (FRED) database. Please see section 3 for a summary of the data.

2.3 Earnest Visualization: Iterative Improvement

2.3.1 Understanding The Data

For the earnest visualization, I want to accurately show whether or not it has become more difficult for the average american to own a home.

We have all the necessary data to answer this question, specifically, for plotting the real wages of the average american against the real home prices.

To eliminate all possible bias, we will use the Real Median Household Income and the Real Residential Property Prices datasets. By using the median of household income, we can avoid the bias of outliers, more accurately showing the average american, and, by using the real prices, we can avoid the bias of inflation.

Let's visualize the bias I am talking about:

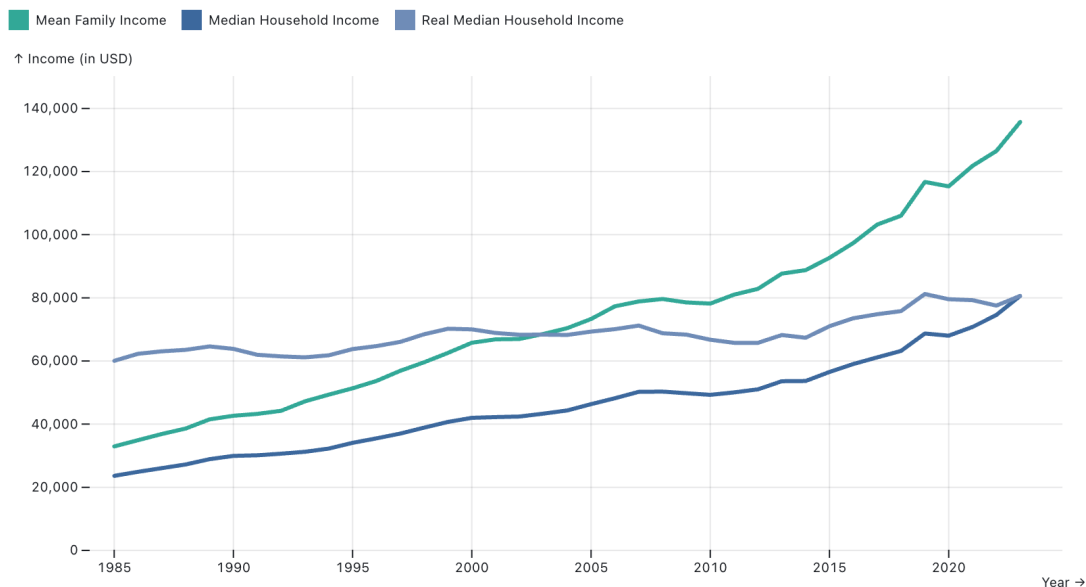


Figure 1: To much of the dismay of my fellow compatriots, the real median household income has been pretty stagnant. However, do realize that we can misrepresent the income of Americans by forgetting this fact.

2.3.2 Iterative Improvement

For the first iteration, I plotted the Real Median Household Income against the Real Residential Property Prices.

Figure 2 does not take inflation into account, so, for our earnest visualization we have two options:

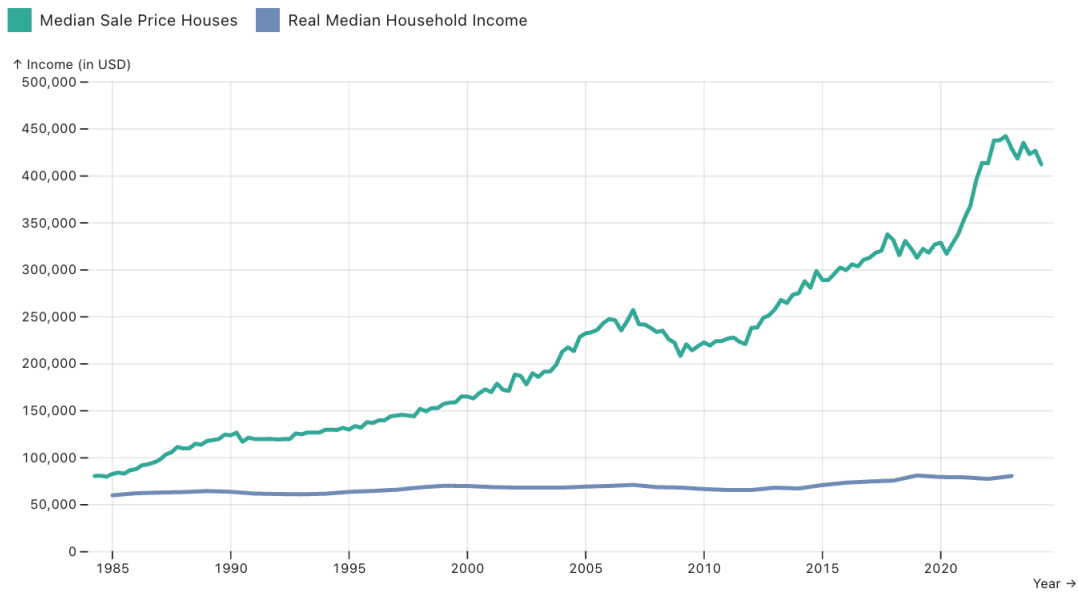


Figure 2: Real Median Household Income vs Residential Property Prices

- Adjust the Real Median Household Income for inflation.
- Utilize Median Household Income without adjusting for inflation, which would represent the price paid for homes in the year of purchase.

Both options are totally fine, but I will go with the first option, since we want to make an appeal to modern individuals.

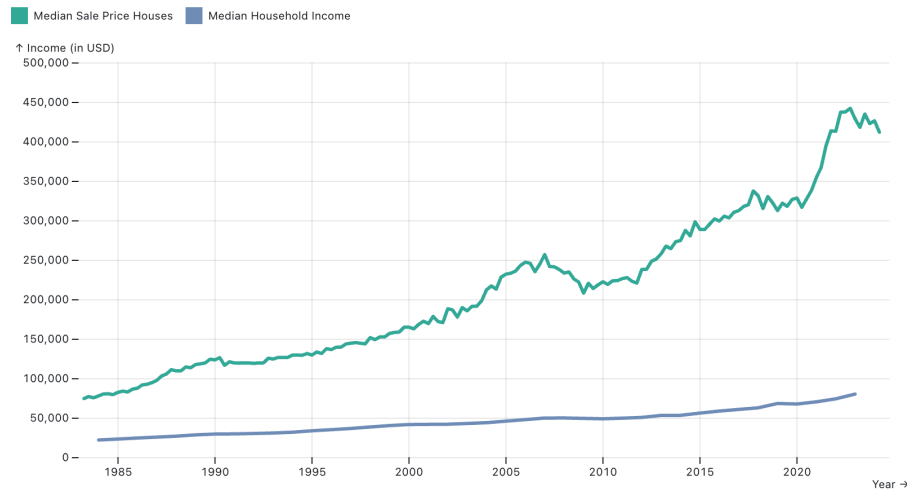


Figure 3: Median Household Income vs Residential Property Prices (No Inflation Adjustment)

In Figure 3 we can see both values without inflation adjustments.

In Figure 4, we can see the values with inflation adjustments. This was done by taking the Consumer Price Index for All Urban Consumers dataset and dividing dollar values for both nominal datasets by it, using the year 2000 as the base year.

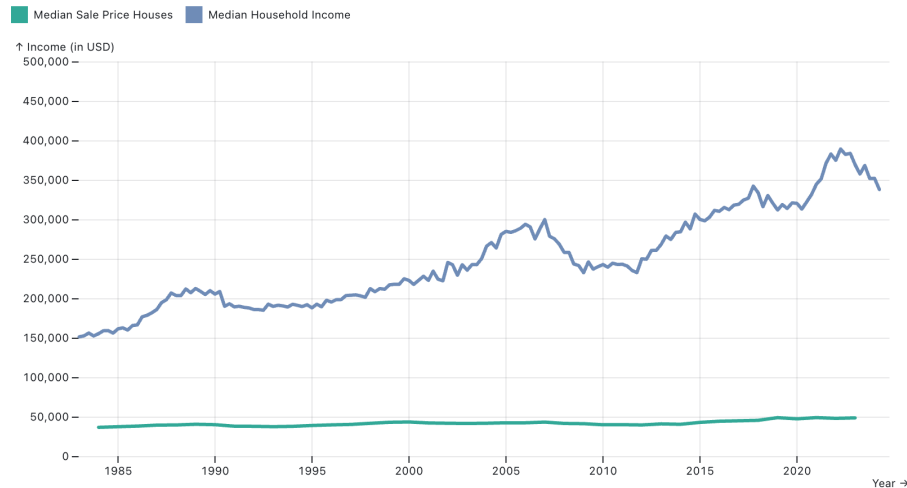


Figure 4: Median Household Income vs Residential Property Prices (No Inflation Adjustment)

Finally, for the earnest visualization, we will add two goodies to improve the visuals:

- A clear title that is not optionated.
- A highlight of recession years to further indicate home price fluctuations.

Real Median Home Sale Price vs Real Median Household Income: US (January 1st 2000 as reference year)

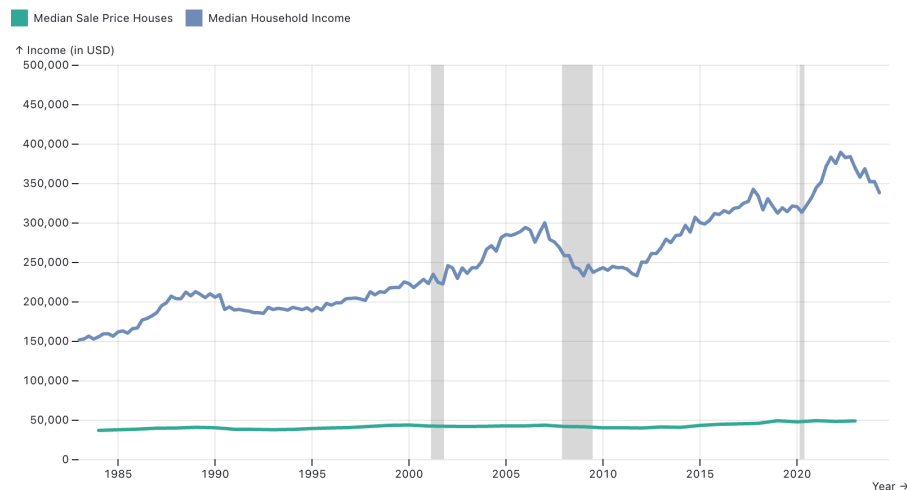


Figure 5: Final Visualization. Beautiful, pristine, and honest.

Hopefully, I have convinced you that there is a disconnect between home prices and income. Please note that the utmost financial care was put into the numbers. The problem is real. Let's see how we can maliciously flip the script.

2.4 misleading Visualization

2.4.1 Iterative Deceit

For the misleading visualization, we will reuse the normalized home price dataset.

However, we will use the Nominal Mean Family Income dataset, for maximum deceit.

Additionally, we will include a non-normalized figure of personal expenditures, to further confuse the viewer.

Hopefully, we can make the viewer believe that the average american is spending more than they are earning, and that that is the reason why they cannot afford a home. By the way, we are intentionally leaving out the rule here, since this is an index, and plotting it correctly would give things away.

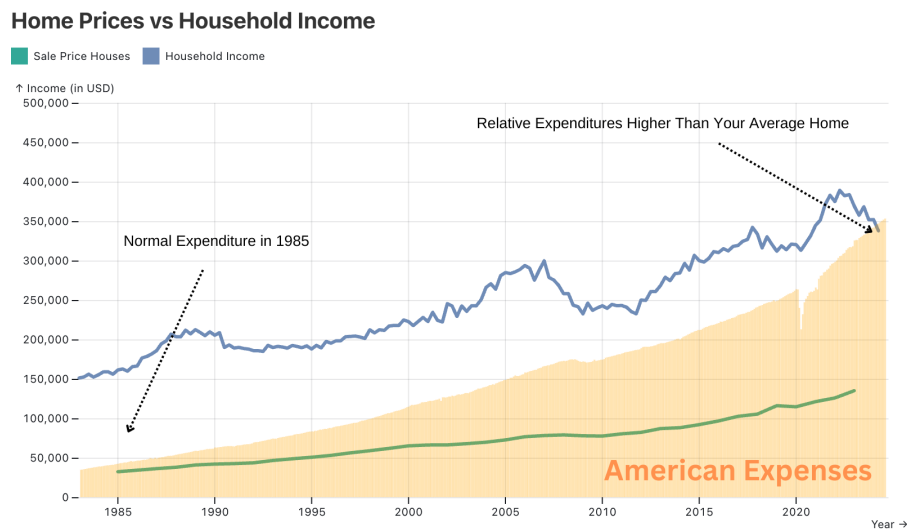


Figure 6: misleading Visualization. The average american is spending more than they are earning.

2.4.2 Effectiveness

I think that this visualization is perfect for confirming an inherent bias in the viewer.

My goal was to show the viewer that the average american has been earning more money, and that, relative to the price of houses, they are more or less the same as they were in the past.

However, by including the personal expenditures, I have made it seem like the average american is spending more than they are earning. Wrongly allowing them to conclude that the problem is with the mentality of young american people, and not with policy.

Of course, these is all bologna. The average american is earning less, and the price of homes has skyrocketed.

3 Apendix

Income, Earnings, and Revenue

- **Mean Family Income**
 - **Series ID:** MAFAINUSA646N
 - **Units:** Current dollars
- **Nominal Median Household Income**
 - **Series ID:** MEHOINUSA646N
 - **Units:** Current dollars
- **Real Median Household Income**
 - **Series ID:** MEHOINUSA672N
 - **Units:** Constant dollars (adjusted for inflation)
- **Income Before Taxes (Wages and Salaries) by Age (from Age 25 to 34)**
 - **Series ID:** CXU900000LB0403M
 - **Units:** Current Dollars

Housing Related

- **Average Sales Price of Houses Sold**
 - **Series ID:** ASPUS
 - **Units:** Current Dollars
- **Median Sales Price of Houses Sold**
 - **Series ID:** MSPUS
 - **Units:** Current Dollars
- **Housing Affordability Index**
 - **Series ID:** FIXHAI
 - **Units:** Index

Personal Consumption

- **Real Personal Consumption Expenditures**
 - **Series ID:** PCEC96
 - **Units:** Billions of Chained 2017 Dollars
- **Personal Consumption Expenditures**
 - **Series ID:** PCE
 - **Units:** Billions of Current Dollars
- **Personal Consumption Expenditures Excluding Food and Energy**
 - **Series ID:** DPCCRC1M027SBEA
 - **Units:** Billions of Current Dollars

Inflation

- **Consumer Price Index for All Urban Consumers: All Items in U.S. City Average**

- **Series ID:** CPIAUCSL
- **Units:** Index 1982-1984=100

References

- [1] *Border Crossing Entry Data*. Accessed: 2023-10-01. 2024. URL: <https://catalog.data.gov/dataset/border-crossing-entry-data-683ae>.
- [2] *Drug Overdose Death Rates by Drug Type, Sex, Age, Race, and Hispanic Origin: United States*. Accessed: 2023-10-01. 2024. URL: <https://catalog.data.gov/dataset/drug-overdose-death-rates-by-drug-type-sex-age-race-and-hispanic-origin-united-states-3f72f>.